2019 Git Contributor Summit – Topics

Elijah Newren

Palantir Technologies

2019-02-05

Setup

- 30-40 people
  - I got to put lots of faces to names
  - Had many top contributors, but definitely not all of them
- Big projects & companies represented:
  - GitHub, GitLab, Atlassian, Gerrit, Microsoft, Google
- Peff’s comment on merge work

Overview

- Fetch response & CDN offloading
- A review of "Big"
- NewHash
- Poll: prove vs. jumbled output
- Poll: where should Git Merge be next year?
- Using mailmap by default in git log?
- Fetch response sideband-all
- Protocol v2 for push
- Partial Clones
- Informal Break discussions
- Breakout Sessions
- CI

Fetch response & CDN offloading

- allows resumable cloning
- does load balancing
- gets the static part of history (e.g. until a week ago) from CDN, and last bits from "main" server
- questions about whether to do multiple bits offloaded (e.g. almost full clone, only stuff from last month, etc.); can server keep track of manifest and direct client to necessary subset of pack on a CDN?

A review of "Big"

- references, history, wide-checkout, deep-checkout, lots to gc, etc.
- newer stuff: partial clones, worktrees, commit-graph
- originally planned to do a breakout session later, but...
- git-sizer (usage by GitLab vs. GitHub)
- large blobs, partial clones
- partial or hierarchical indexes

NewHash

- sha1 -> sha256
- have sha256 repo locally talking to a server using sha1?
- as of yesterday, binary that can create either sha1 or sha256 repos
- will be using fixed length listing of shas in packfile; if given sha1 is fourth in list, then the corresponding sha256 will be fourth
- next: interoperation; fetch & push coming up next
- done a fair amount of work so moving to a new hash in the future with a different length should be much less work
- no automatic translation of commit messages, but maintenance of dual-mapping of hashes
- (Comments on sha1dc & its performance)
- Submodules is the biggest issue right now

Poll: prove vs. jumbled output

Poll: where should Git Merge be next year?

Using mailmap by default in git log?

- People change names for lots of reasons (including transliteration differences)
- Keep an option to not use mailmap
- People generally positive on the idea

Fetch response sideband-all

- sidebands for progress messages and errors
- sideband currently limited to when sending packfile
- proposal: expand sideband for whole response, not just packfile.
- particularly useful given ideas to do CDN
- also needed for keep-alive messages
- this will be a negotiated new capability (can't do it backward-compatibly)

Protocol v2 for push

- ref advertisement the main issue
- Some want to modify the commit server-side (?!?!?)
  - tweak the commit message somehow
  - rebase-on-push
  - reformat-on-push
- discussion of how to split messages up into sub-commands
- a way to retry pushes without re-pushing everything (e.g. someone else updated the branch, you then re-merged or rebased locally and want to push again, meaning the server already has *most* the objects but just needs a few new ones)

Partial Clones

- doing work to have multiple remotes (related: CDN usage)
- still very tied to having a server around to request additional objects
- we need to have a way to keep upload-pack open and do multiple requests
- has rudimentary ability to filter trees
  - Matthew Devore doing some work in this area right now, but it appears to be based on depth rather than width?
- connection with sparse checkout is kind of hacky right now
- there are reachability enforcement issues in V2, which becomes even more of an issue with partial clones (now need to worry about blobs not just commits)
- in a partial clone world, server can't gc
- sidenote: dumb http support
  - no major hosting provider supports it
  - some people like it due to resumability (e.g. Joey Hess & git annex)
  - cgit (repo browser) provides dumb http support natively

CI

- Folks aren't all that happy with the Travis setup
- Dscho has a lot of machinery built up around Azure Pipelines
- PRs to github.com:gitgitgadget/git will automatically be built on Windows, MacOS X, and linux
- Interest in getting emails for failures that their topic branch caused (note: get topic author from tip commit author if not Junio)
- This may be able to move to github.com:git/git after Dscho's patches merge down

Breakout Sessions

Four simultaneous breakout sessions were held:

- merging
- GSoC
- structured logging
- windows big files

Breakout Session: merging

I led the breakout session on merging

- merge-recursive rewrite
  - questions and basic explanation of how the algorithm works
  - want incremental updates on merge-recursive rewrite
  - make merge-recursive code part of libgit.a ?
  - people are very happy about idea to not touch the working tree
- make rebase --merge the default
  - use performance tests to see how well it compares (p3400-rebase.sh)
  - may later also reimplement the am-specific flags on top of interactive
- make use of best merge bases in more places (e.g. git diff A...B uses a suboptimal one)
- rebase --rebase-merges; more thorough reimplementation ideas

Informal Break discussions

- saving incomplete merge state
- checkout overlay
- filter-repo